Parameter → population

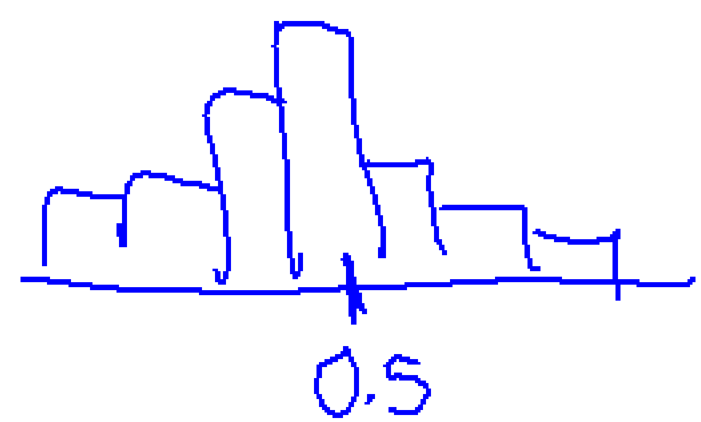
Statistic → sample

Med = res.

* $\bar{x}, \mu, \sigma \Rightarrow$ non-resistant
affected by outliers

Unbiased estimator =
when the center of sampl. distr.
is = to pop. param.

Empirical Rule =
68% - 95% - 99.7%



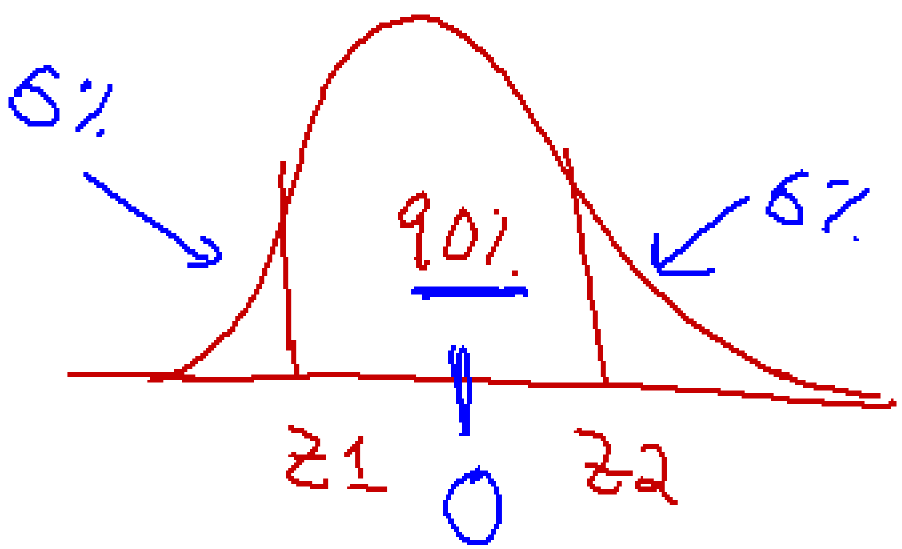
Standard Normal Curve → $N(0, 1)$

z-scores

Standardizing:

$$z = \frac{X-\mu}{\sigma}$$ = how many σ the obs was from (+ or −) its mean.

z-score 90% btw. $z=-1.4$ $z=-0.9$



$z_1 = \text{invnorm}(0.05, 0, 1) = -1.645$

$z_2 = \text{invnorm}(0.95, 0, 1) = 1.645$

CH. 1-5 = Exploratory Data Analysis

* CH. 6-10 = Formal Statistical Inference

Inference = drawing conclusions about a pop. from data/stats.

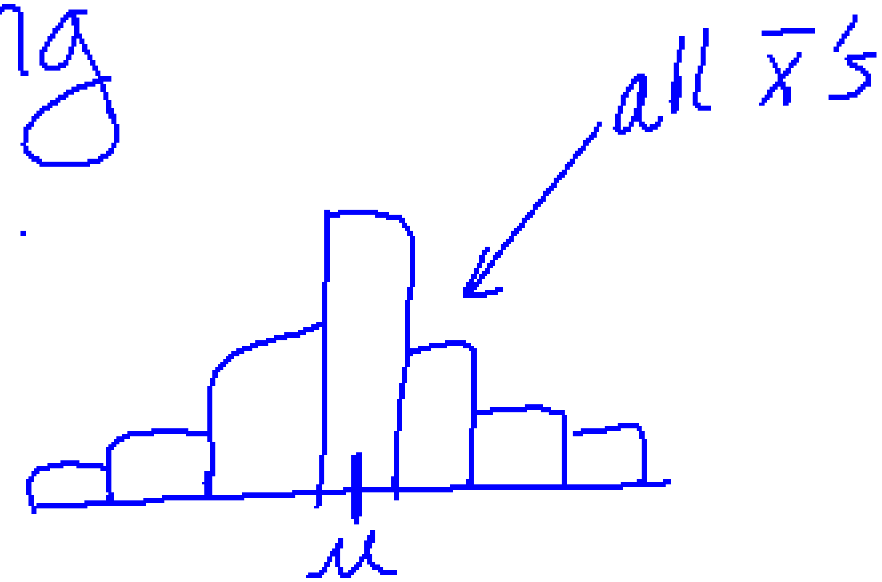
* Formal Stat. Inf = same but with a known degree of confidence (%)

1. Confidence Intervals
   - estimating a population parameter (like μ or p)
     unknown

2. Tests of Significance
   aka: Hypothesis tests
   the evidence for a claim about pop. param (like $\mu = 10$ or $p = 0.30$)

samples
$\bar{X} = 20$
$\bar{X} = 9.8$
prob. calc.

- based on the sampling distrib. of a statistic.



- probabilities of what would happen if we did repeated samplings

$\mu = ?$

$\mu = 10$ $\quad P(\bar{x} = 20 \mid \mu = 10) =$ small

$\bar{x} = 20$

<u>assumptions</u> - random sample (SRS = simple rand samp)

μ - good exp. conditions

CH. 6 ⇒ know σ (pop. std. dev.)

CH. 6 - 10 in general

- take samples
- find statistics ($\bar{x}$, $\hat{p}$, s)
- estimate parameters (μ, p, σ)

* make concl. about param. with a known degree of conf.

Confidence Intervals

Pres. poll: 46% ± 3% margin of error

$n = 1200$

42 – 48%

Form:

estimate ± margin of error

μ

$\bar{x}$ = estimate → μ

↓ m.o.e. → how accurate you believe your guess is.

$45\% \pm 3\%$ 60% conf.

$48\% \pm 10\%$ 90% conf.

↑ conf. level – how conf. we are that our interval contains/catches the true param

1) Interval (a, b)

2) conf. level

in repeated samplings,

the interval contains param.

μ • sampling distributions

• pop $\sim N(\mu, \sigma)$

• sample $\sim N(\mu, \sigma/\sqrt{n})$

Formula:

means

$$\bar{X} \pm (z^*)(\sigma/\sqrt{n})$$

prop:

$$\hat{p} \pm (\quad)\left(\sqrt{\frac{p(1-p)}{n}}\right)$$

Generic:

statistic $\pm$ (critical value)(std. dev. of statistic)

$\bar{x}$ = estimate = statistic

$z^* \sigma/\sqrt{n}$ = margin of error

z^* = critical value

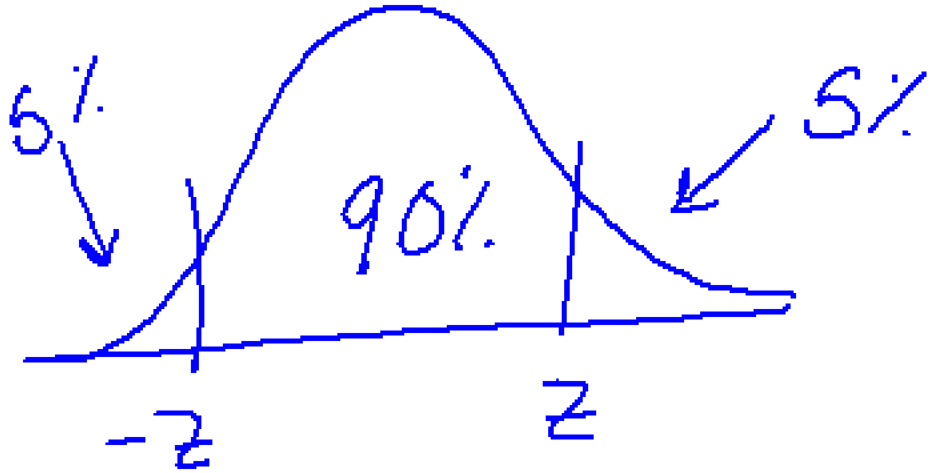
= the z-score that has ___ %

of data in btw $\pm z$.

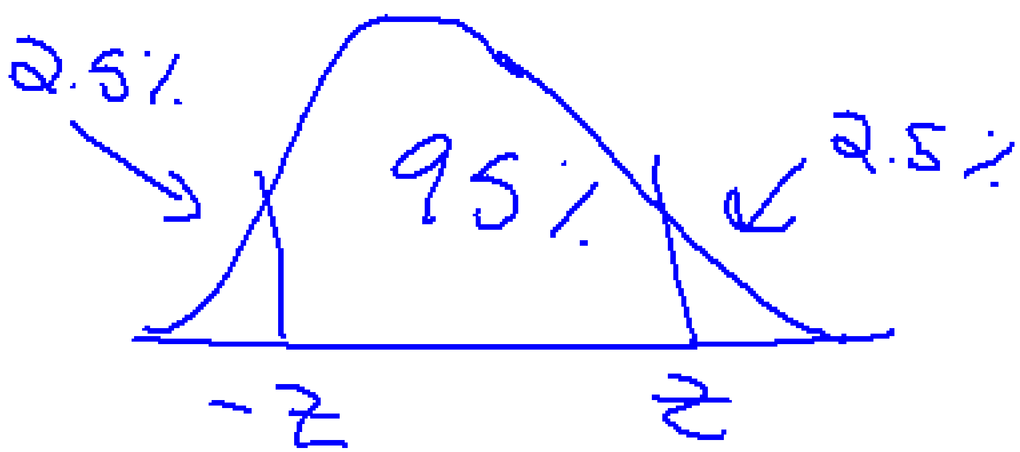
level of conf.

90%, 95%, 99%

90% conf.



$$-z = \text{invnorm}(0.05, 0, 1) = -1.645$$



$$-z = \text{invnorm}(0.025, 0, 1) = 1.96$$

	90%	95%	99%
z^*	1.645	1.96	2.576

Ex 1

95% conf.

$\bar{x} = 25.7$

$\sigma = 3.4$

$n = 100$

$$\bar{x} \pm z^*\left(\frac{\sigma}{\sqrt{n}}\right) \qquad (a, b)$$

$$25.7 \pm (1.96)\left(\frac{3.4}{\sqrt{100}}\right) = (25.0336, 26.3664)$$

Ex. 2

$\sigma = 3$

90%

$n = 4$

$\bar{x} = 190.5$

We are 90% confident that Tim's average wt is btw 188.0325 and 192.9675 lbs.

$$\bar{x} \pm z^* \left(\frac{\sigma}{\sqrt{n}}\right) =$$

$$190.5 \pm 1.645\left(\frac{3}{\sqrt{4}}\right) = (188.0325, 192.9675)$$

$$190.5 \pm 2.4675$$

$90\% = (188.03, 192.97)$

$95\% = (187.56, 193.44)$

$99\% = (186.64, 194.36)$

as conf. level $\uparrow$, interval gets wider.

• interpretation is...

a full sentence in context of problem.

(a, b)

Form:

We are ____ % conf that the mean of ______ is btw ____ and ____ units.

- margin of error:

  $$z^*(\sigma/\sqrt{n})$$

- LOW!
- 3 things to reduce:
  - * lower conf. level
  - * ↑ n
  - ↓ σ (can't actually do)